

Genomic and epigenomic insight into giga-chromosome architecture and adaptive evolution of Royal Lily (*Lilium regale*)

Corresponding Author: Professor Li Guo

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

It is challenging to assemble large genomes and therefore our understanding of their architecture and evolution remains limited. Sun et al. present the chromosome-level reference genome of *Lilium regale*, which represents the largest plant genome sequenced to date. Further analyses reveal that epigenetic regulation and TE expansion contributed to the genome enlargement. Evolutionary analyses provide some insights into adaptation to the environment. The 3D chromatin structure itself is also interesting. In general, this manuscript is well-written, and this genome could be very useful for evolutionary studies and future lily breeding. I have some comments list as follows to improve this manuscript for publication.

Major points:

1. Some results were descriptive, which could not make a strong conclusion. For example, the authors found that DNA methylation was associated with long genes and introns with TEs, and DNA methylation may involve in alternative splicing, however, whether DNA methylation played a regulatory role in these processes was not well supported by current data.
2. In relate to the above comment, some conclusion in the model (Fig.6) is misleading, thus I suggest removing the Fig.6.
3. This study presents a genome with giga-chromosomes. It would be interesting to compare and discuss the common and/or specific genomic features with other recent published genomes with giga chromosomes, such as *L. davidii* var. *unicolor*.
4. The figure legends should be substantially expanded. Many details were missing, which are crucial for understanding these figures.

Minor points:

1. Table 1: the unit format for comparison should be consistent between the two genomes.
2. Fig 1c: replace a-h to A-H.
3. Fig 1e: what are the color and lines represent for?
4. Fig 1g: what the number and line represent for? The 1st pair of chromosomes seems different from other others, and the difference should be discussed.
5. The y axis title for Fig. 2f, 3f, 5f are missing.
6. The values should be defined in figure legends. For example, "DNA methylation" in Fig.2g, "methylation" in Fig 3e, 4e, 5e, "density" in Fig 4c

Reviewer #2

(Remarks to the Author)

The manuscript presents a high-quality chromosome-level assembly of the royal lily (*Lilium regale*) genome, which at 35+ Gb is one of the largest plant genomes ever sequenced. The study uses PacBio HiFi long reads and Hi-C sequencing to assemble and analyze this complex giga-genome. It highlights that genome expansion in *L. regale* is primarily driven by the proliferation of transposable elements (TEs), in particular long terminal repeat retrotransposons (LTR-RTs). Furthermore, the potential mechanisms of TE activity and gene expression regulation are investigated using epigenomic and transcriptomic approaches. Several valuable findings emerge from the study: (a) TEs are regulated by DNA methylation and histone

modifications, contributing to genome stability; (b) TE insertions affect splicing complexity, influencing intron length and gene expression patterns, with potential consequences for adaptive evolution; (c) high-order chromatin structures (TADs) facilitate genome organization and transcriptional regulation. All of these findings are based on robust data sets analyzed using appropriate approaches. Furthermore, the assembled genome can serve as a vital resource for genetics, breeding and conservation biology, as well as for further studies focused on the evolution of large genomes and their regulatory mechanisms.

Although I'm generally positive about the results presented in the manuscript, I also have some concerns. These are mainly related to the way these findings are presented and communicated. In addition, some of the results presented could benefit from additional analyses or inclusion of additional data. My main concerns are:

- 1) The results section is very long and detailed. Authors should find a way to condense it.
- 2) The presentation of the figures is inadequate. While the figures themselves are nice, the figure captions are short and do not do an adequate job of explanation of the figures. This applies to the figures in the main text as well as the extended data figures.
- 3) The manuscript identifies several whole genome duplication (WGD) events based on the identification of syntenic blocks and subsequent estimates of synonymous substitutions between paralogs. This approach, however, is poorly described. The underlying principle of the analysis, the meaning of the results, and the rationale for their interpretation should be explained. Otherwise, readers may be confused about what the different Ks peaks mean or how the number of WGD events for royal lily can be inferred from its syntenic ratios to other genomes.
- 4) The comparative genomics and phylogenomics would benefit greatly if the recently assembled *Lilium davidii* genome were included in the analyses. Especially in terms of gene and intron length or identification of genes unique to *L. regale*. In this context, I'm not sure which general patterns (gene length increase due to increased intron size caused by TE insertions, expression and alternative splicing of the ultra-long genes, WGD events) were already described based on the *L. davidii* assembly and which ones are new and based on the presented data. I don't understand why the *L. davidii* assembly was not included in the comparative analyses or used to discuss the current results.

There are a few minor comments that need clarification beyond the more conceptual issues mentioned above:

L89: "estimated genome size of 35.8 Gb..." Different values of genome size are given throughout the manuscript - in Supplementary Table 2 this value is 36.12 Gb, in L102 35.6 Gb. I wonder which value is correct? And was the genome size estimated by an independent method, ideally flow cytometry?

L107: This section of the manuscript is about assembly evaluation and refers to Fig. 1c and Extended data Fig. 1c as proof that there are no misassembled blocks or misplaced contigs. I'm not sure how the referenced circos plots prove this.

L117: "...using RNA-seq.." However, ISOseq data was also used.

L125: References to the *Allium* genomes included in the comparison are missing,0

L127-128: The number of orthogroups composed of single-copy orthologs seems rather low to me. Is there an explanation for this?

L129: Was the phylogenetic tree based on 24 loci (single-copy genes)?

L131-135: Would the number of unique genes be as high if *L. davidii* were included in the analysis?

L172-173: refers to Supplementary Table 2, but the information in the text and that in the table are different; while the text indicates the proportions of LTR retrotransposons of the total TEs, the table indicates the proportions of LTR retrotransposons in the genome. This should be consistent.

L174-176: This sentence is redundant, gives the same information as the above statement.

L177-180: This sentence is unclear, should be reformulated.

L182-183: The text refers to Figure 2b and Extended data Figure 3b when mentioning the number of intact Gypsy and Copia elements, but these figures do not show these numbers. In fact, they show the phylogenetic relationships within the two LTR superfamilies.

L188-189: "...comprising 3.28%, 19.63%, and 49.20% of the total LTR content... (Supplementary Table 9)" Again, the table has different numbers. This should be consistent.

L190-192: The oldest LTR insertion is dated to 6.39 Mya, but this value is not visible as a distinct peak in Fig. 2c, so it has no relevance for genome size expansions due to increased LTR activity.

L210-211: The reported stronger H3K9Me2 enrichment at the termini of the intact LTRs is not obvious from Fig. 2e, which is

referenced to demonstrate it.

L214-215: "However, the truncated LTR resulting from this insertion..." What information could be used to infer the insertion time of a truncated LTR retrotransposon (i.e., an LTR retrotransposon that does not contain terminal repeats)?

L262-263: The reported differences in TE insertions and DNA methylation between introns and exons are not obvious from Figure 3f. The exons and introns should be clearly labeled in the figure.

L296-297: I don't see the described patterns in the referenced figures.

L367-368: "...most of the TPR genes involved in RNA splicing were unique to *L. regale*, suggesting the presence of a unique splicing mechanism in *L. regale*..." I wonder if this mechanism is really unique to the royal lily or is it more genus or family specific? Again, this could be answered by including the *L. davidii* genome in the analysis.

L604: What script was used to identify the solo LTRs?

L612-614: Why were the protein data from *L. davidii* not used for homology-based gene identification along with the others mentioned?

L637-641: divergence time dating was performed, but the results of this analysis are not included in the manuscript.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

My concerns have been properly addressed and this manuscript was comprehensively improved. The Royal Lily genome could be a valuable resource for plant community.

Reviewer #2

(Remarks to the Author)

I have reviewed the revised version of the manuscript. The authors have addressed all the comments I raised in my previous review. They shortened the results section, improved the clarity and informational value of the figure captions, and provided more detailed explanations of some of the results. Moreover, they included additional *Lilium* genomes and reworked the comparative genomic analyses, enabling the distinction of patterns specific to *L. regale* from those shared with other *Lilium* species. I have no further comments on the manuscript and recommend it for publication.

permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Point-to-point response to the reviewer comments.

Reviewer #1 (Remarks to the Author):

It is challenging to assemble large genomes and therefore our understanding of their architecture and evolution remains limited. Sun et al. present the chromosome-level reference genome of *Lilium regale*, which represents the largest plant genome sequenced to date. Further analyses reveal that epigenetic regulation and TE expansion contributed to the genome enlargement. Evolutionary analyses provide some insights into adaptation to the environment. The 3D chromatin structure itself is also interesting. In general, this manuscript is well-written, and this genome could be very useful for evolutionary studies and future lily breeding. I have some comments list as follows to improve this manuscript for publication.

Response: We sincerely thank you for the thorough and insightful evaluation of our manuscript, as well as your constructive suggestions to further enhance the quality of this work. We greatly appreciate your recognition of the significance of the *Lilium regale* genome as a critical resource for evolutionary studies and lily breeding.

Comment 1. Some results were descriptive, which could not make a strong conclusion. For example, the authors found that DNA methylation was associated with long genes and introns with TEs, and DNA methylation may involve in alternative splicing, however, whether DNA methylation played a regulatory role in these processes was not well supported by current data.

Response: Thank you for this valuable comment. Like you said, we did observe a correlation between DNA methylation was associated with long genes and introns with TEs, and DNA methylation may involve in alternative splicing. However, we agree with you that we do not have direct evidence to support a causal relationship between DNA methylation and these biological processes. Therefore, we have revised the manuscript by toning down our use of words in discussion section (changes highlighted in red).

Lines 385-399: "We *proposed* that alternative splicing in *L. regale* genome was likely regulated by two key factors: TE insertion and DNA methylations. *Genes with intronic TE insertions, particularly those involving LTRs and LINEs, displayed high splicing activity, especially for A3 events (Fig. 4c,d).* We also observed that DNA methylation in exons and introns of large genes was distinguishable (*Fig. 3e and Extended Data Fig. 4e*), and genes with low splicing activity were often associated with higher DNA methylation at splice sites. *Previous studies showed that TE insertions can alter splicing*

patterns and even turn off gene expression⁶³, for example through CHG methylation at splice sites in maize⁶⁴. We similarly observed that strong CHG methylation in non-TE regions of introns was associated with low splicing activity in *L. regale* (**Extended Data Fig. 5g**). This suggested that the mechanisms by which splicing was affected by DNA methylation and TE insertions were probably different. Notably, genes with low alternative splicing, are primarily involved in RNA formation and splicing (**Supplementary Fig. 8**), and contained fewer TE insertions thus relatively stable, likely benefiting from rapid TE clearance. Combination with high CHG methylation at non-TE sites also might reduce splicing variants to stabilize gene transcription."

Comment 2. In relate to the above comment, some conclusion in the model (Fig.6) is misleading, thus I suggest removing the Fig.6.

Response: Thank you for this comment. We agree that Figure 6 should be removed and has removed it.

Comment 3. This study presents a genome with giga-chromosomes. It would be interesting to compare and discuss the common and/or specific genomic features with other recent published genomes with giga chromosomes, such as *L. davidii* var. unicolor.

Response: Thank you for your valuable suggestions. Per your suggestion, we have incorporated recently published *Lilium* genomes into our comparative analysis. In 2024, two high-quality *Lilium* genomes (*L. davidii* and *L. sargentiae*) and one genome from the Colchicaceae family (*Gloriosa superba*) were published. Including these genomes in our study provides significant support for understanding *Lilium* evolution and breeding. We analyzed the common and/or specific genomic features of the three *Lilium* species in terms of phylogenetic relationships, gene family evolution, and WGD events (**Fig. 1d, and Extended Data Fig. 2a,d**). At the same time, the characters of assembly quality, TE proportion and gene length has also been added into the result section. The following texts have been revised accordingly:

Phylogenetic analysis incorporating the published lily genomes has been added to the results section.

Lines 121-125: "A phylogenetic tree constructed using MCMCTREE³⁶ with the 370 low-copy orthogroups, with divergence times estimated and calibrated using fossil records from TimeTrees³⁷ (**Extended Data Fig. 2a**), showing that *G. superba* and (*L. davidii*, (*L. regale*, *L. sargentiae*)) diverged at ~95.6 million years ago (Mya). *L. regale* and *L. sargentiae* are more closely related than each with *L. davidii*, consistent with

previous transcriptome-based analysis²⁸."

Gene family evolution analysis has been added to the results section.

Lines 113-120: "Ortholog identification and clustering using Orthofinder³⁵ classified 698,581 genes into 62,451 orthogroups with four categories: multi-copy, single-copy, unique-copy and others (**Fig. 1d**), 370 of which were composed of low-copy orthogroups (1 or 2 genes per orthologroup). In *Lilium*, 30,012 unique genes were identified, primarily associated with transcription regulation and post-transcriptional gene silencing (**Supplementary Fig. 3**). In contrast, *L. regale* had 15,691 unique genes, which were enriched in nucleotide excision repair and mRNA 3'-UTR binding. These genes may have evolved as adaptations to the exceptionally large genomes of *Lilium* (**Supplementary Fig. 4**). "

Line 126-130: "Furthermore CAFE³⁸ analysis revealed significantly expanded orthogroups in *Lilium* primarily involved in terpenoid biosynthesis (**Supplementary Fig. 4a**). *L. regale* expanded orthogroups were also enriched with MAPK signaling pathway and defense responses mediated by cell wall thickening (**Supplementary Fig. 4b**). This reflected the genome evolution in *L. regale* to cope with environmental stress during its independent evolution^{39,40}."

WGD analysis has been added to the results section.

Lines 151-156: "Furthermore, synteny analysis between the two lilies *L. regale* and *L. davidii* revealed strong gene-level collinearity (**Fig. 1g** and **Extended Data Fig. 2d**). Previous studies^{23,28} have shown that *L. davidii* and *L. sargentiae* underwent independent WGD events, suggesting that the three *Lilium* species share a similar WGD history. Together, these analyses provide robust evidence that *Lilium* species has undergone an independent WGD."

The comparison of assembly quality has been added to Table 1.

Lines 932-935:

Table 1: Statistics for genome assembly and annotation of *Lilium regale*.

Statistics	<i>Lilium davidii</i>	<i>Lilium sargentiae</i>	<i>Lilium regale</i> MJ-PKU01
Assembly size (bp)	36,680,000,000	47,092,191,210	35,569,194,365
No. of contigs	13,068	85,041	3,608
No. of scaffolds	3,853	35	1,409

No. of chromosomes	12	12	12
Scaffold N50 (bp)	2,860,000,000	2,845,141,673	2,714,556,596
Contig N50 (bp)	7,720,000	929,352	32,225,586
Total length of pseudomolecules (bp)	-	-	34,433,330,367
GC content (%)	34.94	-	32.26
Repeat content (%)	88.31	82.32	80.06
No. of protein-coding genes	87,501	47,139	82,881
No. of high-confidence protein-coding genes	-	-	67,862
Average gene length (bp)	57,607	42,804	19,940
HiFi mapping rate (%)	-	-	99.55
ONT mapping rate (%)	99.10	-	-
NGS mapping rate (%)	97.80	-	95.44
BUSCO of genome (%)	94.90	90.77	96.16
BUSCO of protein (%)	79.50	88.30	88.20
QV (quality value)	30.18	-	63.80

#The lineage dataset is: embryophyta_odb10 (*L. davidii* and *L. sargentiae*) (Creation date: 2024-01-08, number of genomes: 50, number of BUSCOs: 1614); liliopsida_odb10 (*L. regale*) (Creation date: 2024-01-08, number of genomes: 15, number of BUSCOs: 3236)

The comparison of TE constituents has been added to the results section.

Lines 178-181: "*Despite a Gypsy-rich genome, the numbers of intact Gypsy (34,472) and Copia (35,771) elements were comparable (Fig. 2b and Extended Data Fig. 3b), with Tekay and Athila as the most abundant LTR/Gypsy subfamilies (Fig. 2b), which was consistent with the research in L. davidii and L. sargentiae*^{23,28}."

The comparison of gene length has been added to the results section.

Lines 168-170: "*Repeat annotation in L. regale identified 27.9 Gb repetitive sequences, occupying 80.06% of the genome (Supplementary Table 7), showing their substantial contribution to L. regale genome expansion, similar to L. davidii (88.31%) and L. sargentiae (82.32%)*^{23,28}."

Line 213-217: "*L. regale* gene sizes varied substantially ranging from 75bp to 1.51 Mb (average length: 19,940 bp) (**Table 1 and Supplementary Table 10**), longer than typical plant genes mainly attributed to the presence of long introns (average 4,680 bp) much longer than usual^{23,28,47-49} (**Supplementary Table 10**) but shorter than *L. davidii* and *L. sargentiae*^{23,28}."

References:

Liang Y, Gao Q, Li F, et al. The giant genome of lily provides insights into the hybridization of cultivated lilies. *Nat Commun.* 2025 Jan 2;16(1):45.

Xu S, Chen R, Zhang X, et al. The evolutionary tale of lilies: Giant genomes derived from transposon insertions and polyploidization. *Innovation (Camb).* 2024 Oct 24;5(6):100726.

Comment 4. The figure legends should be substantially expanded. Many details were missing, which are crucial for understanding these figures.

Response: Thank you for this suggestion. Now all the figure legends have been expanded and the new contents are highlighted in red.

Comment 5. Table 1: the unit format for comparison should be consistent between the two genomes.

Response: Thank you for this comment and we agree. The unit formats used for inter-genomic comparisons have been adjusted for **Table 1** (above in Comment 1).

Comment 6. Fig 1c: replace a-h to A-H.

Response: Done.

Comment 7. Fig 1e: what are the color and lines represent for?

Response: Thank you for the question. The blue boxes represent syntenic blocks in *Lilium regale* per *Dioscorea alata* block. For example, in the "Chr1 of *L. regale*" column, four blue boxes are connected by blue lines, indicating that these selected blocks are orthologous to "chr1 of *L. regale*.", suggesting a 4:4 syntenic relationship between the two species. Conversely, the red boxes and lines denote the syntenic blocks of *D. alata* per *L. regale* block, similarly showing a 4:4 syntenic relationship between the two species. We apologize for the unclear description in the first place, and have revised the Fig. 1e legend as below.

Fig. 1e "*Dot plot showing the syntenic relationship of homoeologous chromosomes*

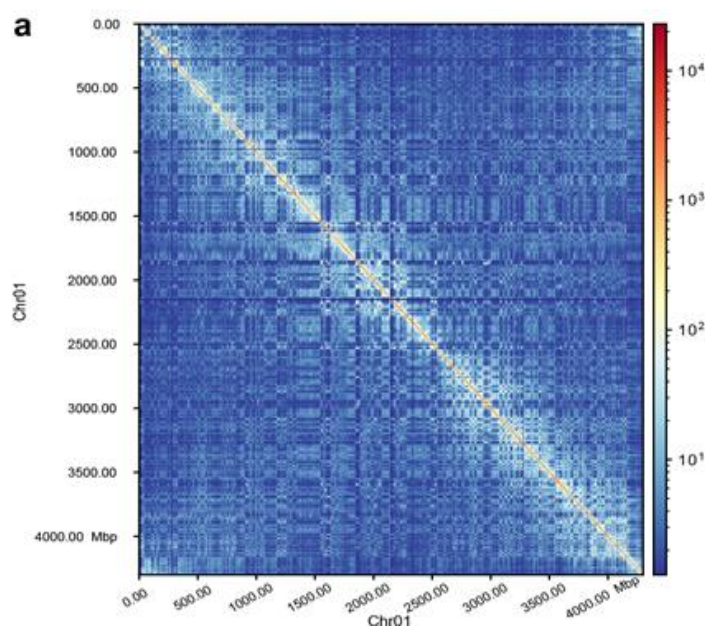
between *L. regale* and *D. alata*. Representative synteny blocks signaling the whole genome duplication events are color highlighted. The blue boxes represent homologous segments of *L. regale* in *D. alata*. Meanwhile, homologous segments from the same homologous segment sets are linked by blue lines. Similarly, the red boxes and lines denote homologous segments of *D. alata* in *L. regale*. "

Comment 8. Fig 1g: what the number and line represent for? The 1st pair of chromosomes seems different from other others, and the difference should be discussed.

Response: Thank you for the question. Numbers represent chromosome IDs, and the lines indicate syntenic blocks, which are conserved genomic regions with shared gene order. We have revised the Fig. 1g legend as follows:

Fig. 1g " Syntenic comparison of *L. regale* with *L. davidii*. Numbers represent chromosome IDs, and the lines indicate syntenic blocks, which are conserved genomic regions with shared gene order."

As for the difference in Chr1, it likely comes from genomic difference between species. To verify whether this is an assembly error in our *L. regale* genome, we examined the Hi-C maps of *L. regale* for Chr1 (Supplementary Figure 1a), which did not reveal any large-scale assembly errors, suggesting that our *L. regale* assembly is accurate. Therefore, the observed rearrangement between the two genomes is likely due to genomic variation in the two species.



Supplementary Figure 1a. Hi-C interaction heatmaps for pseudomolecules Chr1.

Comment 9. The y axis title for Fig. 2f, 3f, 5f are missing.

Response: Thank you for this comment. The y axis title for Fig. 2f, 3f, 5f have been added and figure legends have been updated accordingly, as shown below.

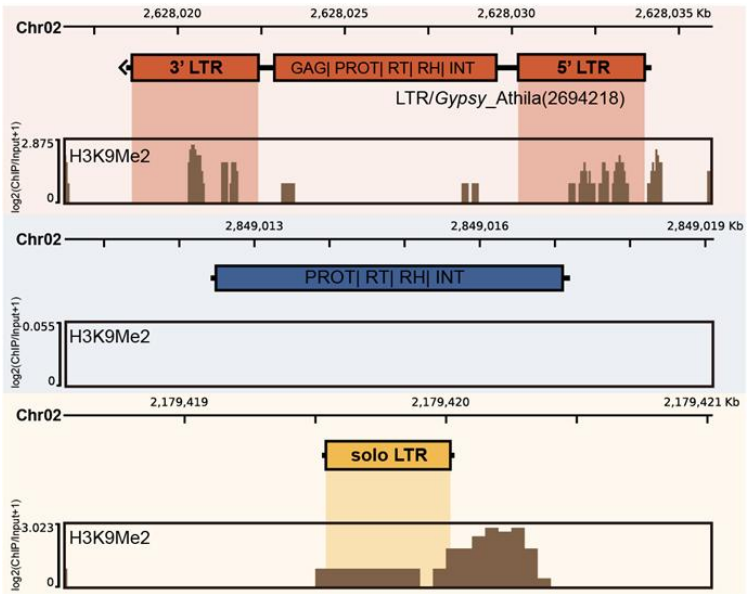


Figure 2f. The intact LTR/Gypsy in Chr02 and its truncated LTR and solo LTR as an example. H3K9Me2 is enriched on LTR repeat region.

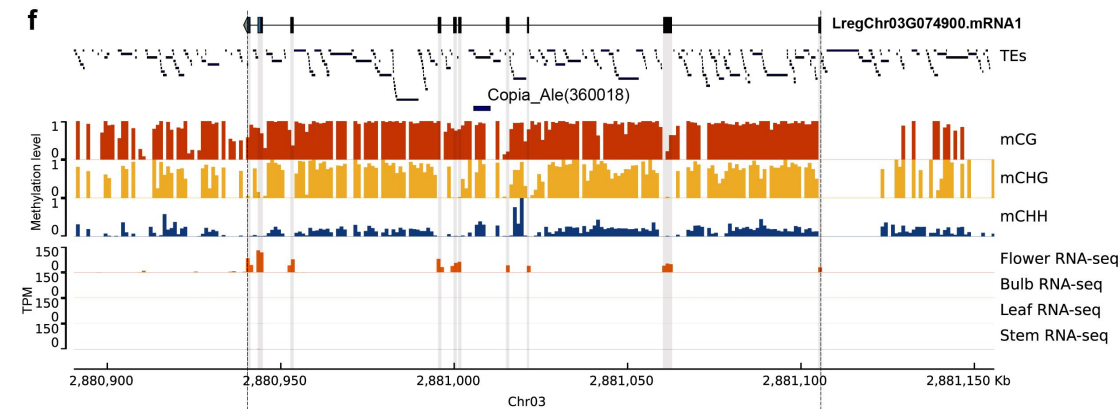


Figure 3f. with many TE insertions, *LregChr03G072890* expression under the control of methylation. The gene structure, TE insertions, intact LTR with its insertion time (year), methylation levels (CG, red; CHG, yellow; and CHH, blue; methylation level represent the proportion of 5mC at the given site), gene expression in four tissues (flower, orange; bulb, canary; leaf, cyan; and stem, purple) were plotted from the top to bottom successively. The shaded gray areas represent exon regions.

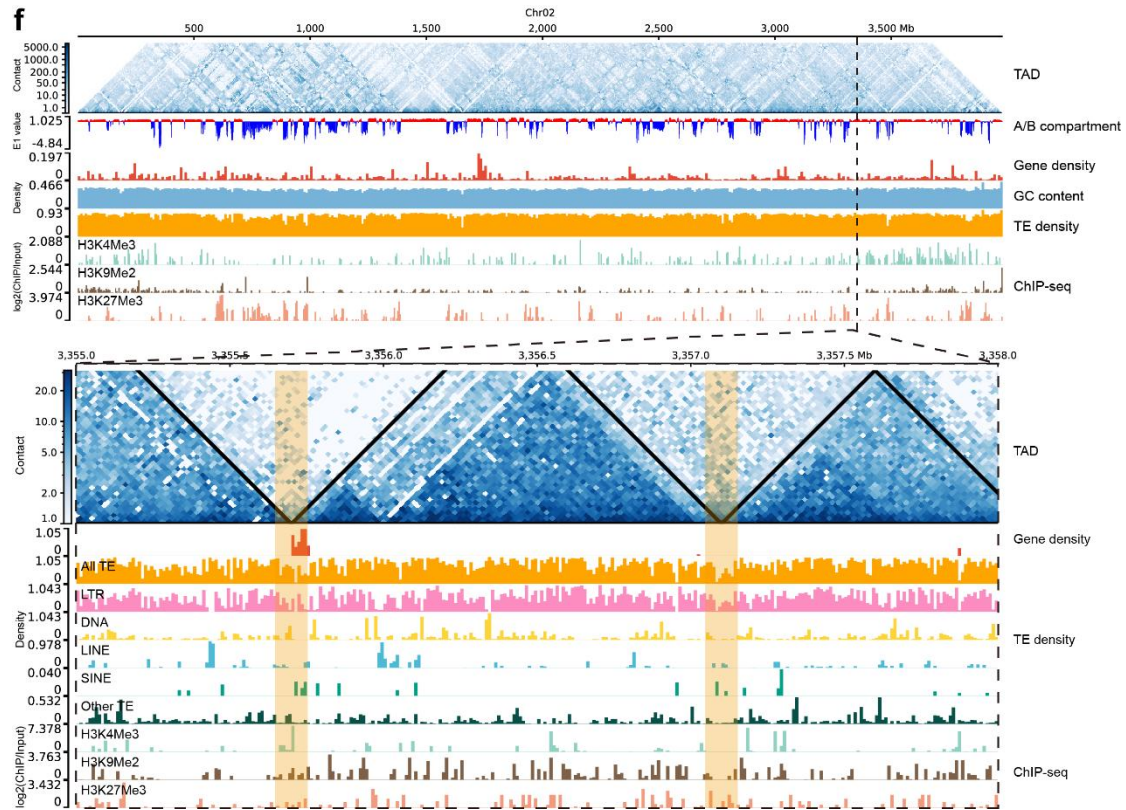


Figure 5f. Characteristics of the region on Chr02 as an example, which had SINEs and H3K9Me2 enrichment on TAD boundaries (Highlighted). The TAD domains (black triangles), A/B compartment (red, A compartment; blue, B compartment), gene density (red), GC content (%), TE density (orange, all TE; LTR, pink; DNA, gold; LINE, skyblue; SINE, green; other TE, darkgreen), ChIP-seq (cyan, H3K4Me3; brown, H3K9Me2; flesh, H3K27Me3) were plotted from the top to bottom successively.

Comment 10. The values should be defined in figure legends. For example, “DNA methylation” in Fig.2g, “methylation” in Fig 3e, 4e, 5e, “density” in Fig 4c.

Response: Thank you for this great suggestion. "Methylation Level" represents the proportion of 5mC at the given site, while "density" denotes the TE density (the proportion of TE length to the bases with non-zero coverage in the region) in the gene region and its flanking areas.

The related definition has also been added into corresponding figure legends highlighted in red as shown below. The calculation methods are described in more details in the Materials and Methods section.

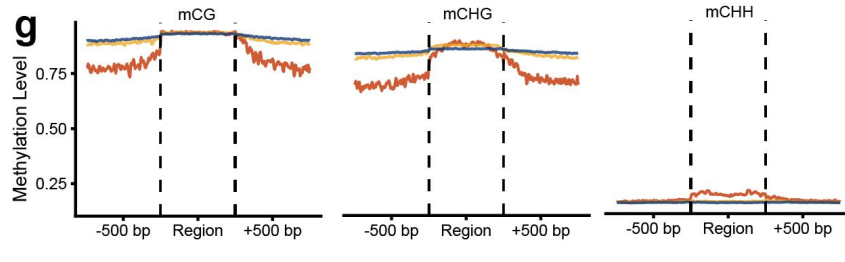


Figure 2g. The DNA methylation ratio (The proportion of 5mC at the given site) on different types of LTRs.

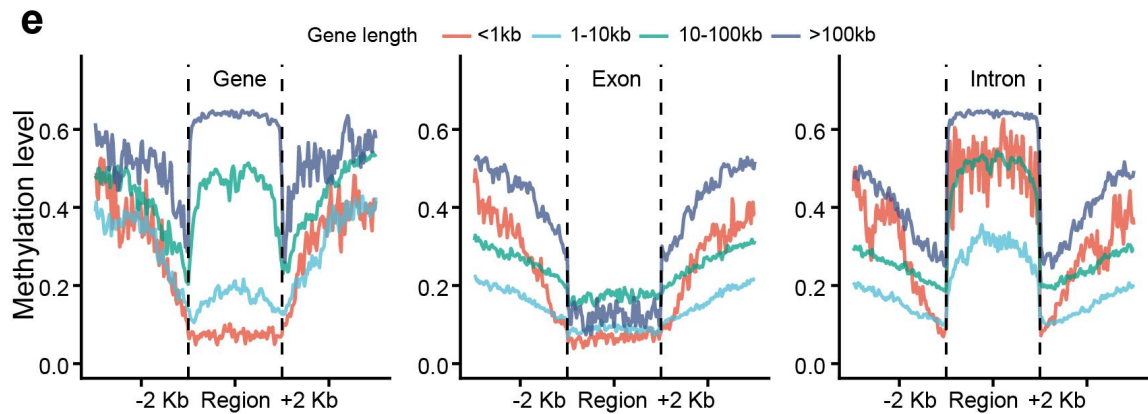


Figure 3e. The DNA Methylation ratio (The proportion of 5mC at the given site) of gene bodies, exons and introns between their upstream and downstream (± 2 Kb) in flower tissue.

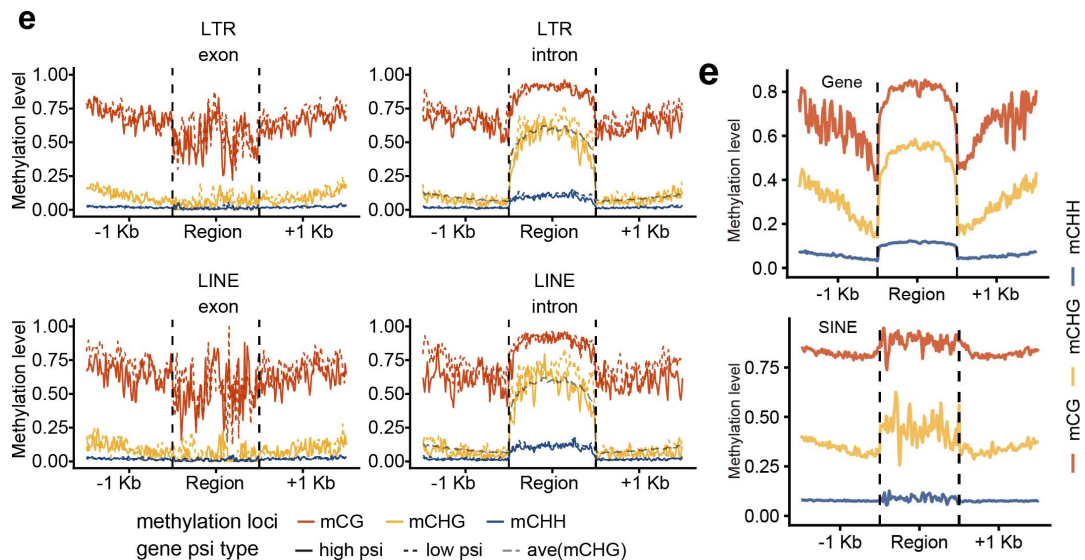


Figure 4e(left) and 5e(right). 4e, The DNA methylation level (CG, red; CHG, yellow; CHH, blue; methylation level represents the proportion of 5mC at the given site) of the gene bodies, upstream and downstream of genes with LTR/LINE insertions (colored line,

HPGs; colored dash, LPGs; grey dash, average of all genes at mCHG). **5e**, The DNA methylation level (CG, red; CHG, yellow; CHH, blue; methylation level represents the proportion of 5mC at the given site) between upstream and downstream of Gene and SINE on TAD boundaries.

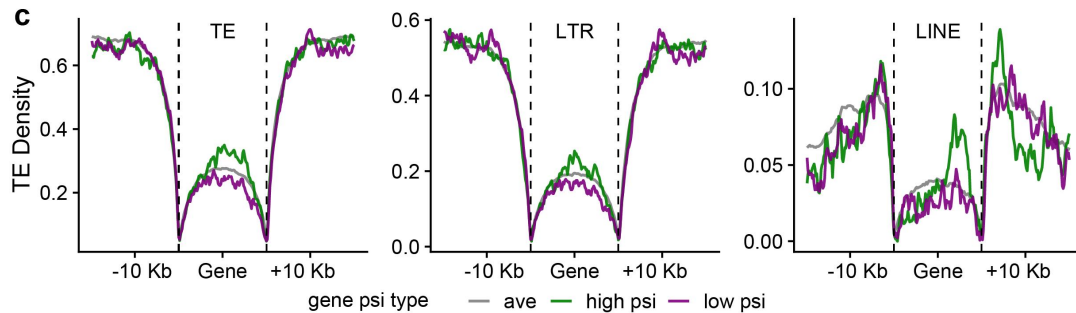


Figure 4c. The TE density (the proportion of TE length to the bases with non-zero coverage in the region) in gene bodies, upstream and downstream of HPG (high PSI gene, green), LPG (low PSI gene, purple) and average of all identified genes by PacBio Iso-seq (grey).

Reviewer #2 (Remarks to the Author)

The manuscript presents a high-quality chromosome-level assembly of the royal lily (*Lilium regale*) genome, which at 35+ Gb is one of the largest plant genomes ever sequenced. The study uses PacBio HiFi long reads and Hi-C sequencing to assemble and analyze this complex giga-genome. It highlights that genome expansion in *L. regale* is primarily driven by the proliferation of transposable elements (TEs), in particular long terminal repeat retrotransposons (LTR-RTs). Furthermore, the potential mechanisms of TE activity and gene expression regulation are investigated using epigenomic and transcriptomic approaches. Several valuable findings emerge from the study: (a) TEs are regulated by DNA methylation and histone modifications, contributing to genome stability; (b) TE insertions affect splicing complexity, influencing intron length and gene expression patterns, with potential consequences for adaptive evolution; (c) high-order chromatin structures (TADs) facilitate genome organization and transcriptional regulation. All of these findings are based on robust data sets analyzed using appropriate approaches. Furthermore, the assembled genome can serve as a vital resource for genetics, breeding and conservation biology, as well as for further studies focused on the evolution of large genomes and their regulatory mechanisms.

Although I'm generally positive about the results presented in the manuscript, I also have some concerns. These are mainly related to the way these findings are presented

and communicated. In addition, some of the results presented could benefit from additional analyses or inclusion of additional data.

Response: We sincerely thank your comprehensive and thoughtful evaluation of our work, as well as your recognition of the technical and biological significance of the *Lilium regale* genome assembly. We also appreciate your positive assessment of the robustness of our datasets, analytical approaches, and the potential utility of this resource for evolutionary genomics, breeding, and conservation biology. To address your concerns, we have improved the manuscript by strengthening the clarity of presentation and expanding certain analyses as explained below.

Comment 1. My main concerns are: The results section is very long and detailed. Authors should find a way to condense it.

Response: Thank you for the suggestion. We have revised the result section by condensing the content and making it more concise, reducing 3784 words to 3311 words.

Comment 2. The presentation of the figures is inadequate. While the figures themselves are nice, the figure captions are short and do not do an adequate job of explanation of the figures. This applies to the figures in the main text as well as the extended data figures.

Response: We apologize for the insufficient information in figure legends. We have improved the figure legends by adding more detailed explanation to each figure, which are highlighted the additional contents in red text (see manuscript text). Thank you for the suggestion.

Comment 3. The manuscript identifies several whole genome duplication (WGD) events based on the identification of syntenic blocks and subsequent estimates of synonymous substitutions between paralogs. This approach, however, is poorly described. The underlying principle of the analysis, the meaning of the results, and the rationale for their interpretation should be explained. Otherwise, readers may be confused about what the different Ks peaks mean or how the number of WGD events for royal lily can be inferred from its syntenic ratios to other genomes.

Response: Thank you for your suggestion. We have added more descriptions about the evolutionary genomic analysis in both the result and method sections. The revisions are shown below:

Lines 133-156: "Additionally, monocots such as rice, maize, and banana have undergone independent WGDs after the τ event⁴³. Interspecies comparison showed a 4:4 synteny ratio (**Fig. 1e**) between *L. regale* and *Dioscorea alata*, a species undergoing two WGD events (the τ event and another independent WGD)⁴⁴, suggesting that *L. regale* also undergone two WGD events. Furthermore, a 4:1 synteny ratio between *L. regale* and the ancestral monocot karyotype⁴⁵ (**Extended Data Fig. 2b**) supported the hypothesis that *L. regale* has undergone two additional WGDs since its divergence from the monocot ancestor.

We then examined the genome duplication history of *L. regale* by analyzing the distribution of synonymous substitution rates (K_s) for syntenic gene pairs identified in synteny analysis (**Fig. 1f**). K_s peaks correspond to genome duplication events (self-synteny) or species-divergence (interspecies synteny). For *L. regale* syntenic paralogs, we observed three K_s peaks (**Fig. 1f**). The first peak ($K_s \approx 1.5$) shared with *D. alata* represented the τ WGD (~125Mya) widely present in monocots including *L. regale* and *D. alata*. The second peak ($K_s \approx 1.09$) represented a WGD event in *L. regale* likely occurring at a similar time (~90.1 Mya) with the second WGD in *D. alata* judging by the two close peaks (**Fig. 1f**). Because the K_s for the two peaks were both smaller than the K_s for divergence peak of *L. regale* and *D. alata*, this second WGD should occur after the two species diverged, suggesting it being an independent WGD for both species (**Extended Data Fig. 2a**). Furthermore, synteny analysis between the two lilies *L. regale* and *L. davidii* revealed strong gene-level collinearity (**Fig. 1g** and **Extended Data Fig. 2d**). Previous studies^{23,28} have shown that *L. davidii* and *L. sargentiae* underwent independent WGD events, suggesting that the three *Lilium* species share a similar WGD history. Together, these analyses provide robust evidence that *Lilium* species has undergone an independent WGD."

Lines 591-613: " *L. regale* and AMK exhibit extensive and well-maintained collinear blocks, with a syntenic block ratio of 4:1 (*L. regale*:AMK), indicating that *L. regale* has undergone WGD. Similarly, *L. regale* and *D. alata* also share extensive and well-maintained collinear blocks, with a ratio of 4:4, further supporting the occurrence of WGD events in *L. regale*. We calculated the K_s values for each gene pair in the aligned blocks obtained using WGD1 (v0.4.7) and plotted the distribution of all K_s values to infer WGD events in *L. regale*'s evolutionary history. Three K_s peaks were observed in *L. regale*, indicating multiple WGD events. Theoretically, each peak represents a WGD event. However, frequent recent gene duplications can also generate K_s peaks. To verify this, we mapped the gene pairs associated with the recent K_s peak

onto chromosomes. These genes were not evenly distributed along the chromosomes, suggesting that the recent *Ks* peak resulted from random gene duplications rather than a true WGD event. The remaining two *Ks* peaks correspond to the τ event shared by monocots and a lineage-specific WGD event in *Lilium*, confirming that *L. regale* has experienced at least two WGD events in its evolutionary history.

To estimate the timing of WGD in *L. regale*, we used the "-ks" parameter in WGD_i to calculate the *Ks* values of syntenic block genes within *L. regale* and between *L. regale* and *D. alata*. We then applied the "-kp" parameter in WGD_i for Gaussian fitting of the *Ks* values, identifying the components associated with WGD peaks and calculating their mean *Ks* values and standard deviations. Using an estimated divergence time of 119.9 Mya and the average *Ks* value (1.45) of syntenic gene pairs between *L. regale* and *D. alata*, we estimated the evolutionary rate of *L. regale*. The synonymous substitution rate per site per million years (*r*) was calculated as $6.046e^{-4}$ ($r = Ks / 2T$). Finally, we applied this *r* value and the mean WGD *Ks* value to estimate the WGD timing in *L. regale* using the formula $T = Ks / 2r$."

Comment 4. The comparative genomics and phylogenomics would benefit greatly if the recently assembled *Lilium davidii* genome were included in the analyses. Especially in terms of gene and intron length or identification of genes unique to *L. regale*. In this context, I'm not sure which general patterns (gene length increase due to increased intron size caused by TE insertions, expression and alternative splicing of the ultra-long genes, WGD events) were already described based on the *L. davidii* assembly and which ones are new and based on the presented data. I don't understand why the *L. davidii* assembly was not included in the comparative analyses or used to discuss the current results.

Response: Thank you for your valuable suggestions. Per your suggestion, we have incorporated recently published *Lilium* genomes into our comparative analysis. In 2024, two *Lilium* genomes (*L. davidii* and *L. sargentiae*) and one (*Gloriosa superba*) from the Colchicaceae family were published. Comparing our *L. regale* genome with these genomes helps understanding *Lilium* evolution and breeding.

Although the three studies all show the impact of transposable elements (TEs) on Lily genome size, the two studies on *L. davidii* and *L. sargentiae*. They only infer that LTRs contribute to genome expansion based on their total length and proportion in the genome. In our study of *L. regale*, we further elucidate the mechanism by which LTRs promote genome expansion, demonstrating that their persistence is due to epigenetic

modifications preventing timely removal. Additionally, we conducted a detailed analysis of the impact of TEs on gene length and revealed that the formation of ultra-long genes is primarily driven by the presence of ultra-long introns, which are mainly caused by LTR insertions. Furthermore, alternative splicing has not been analyzed in previous *Lilium* studies. Therefore, our findings contain novel insights not revealed by previous studies. We have included this analysis in the discussion (**Lines 385-404**:

*"We **proposed** that alternative splicing in *L. regale* genome was likely regulated by two **key** factors: TE insertion and DNA methylations. **Genes with intronic TE insertions, particularly those involving LTRs and LINEs, displayed high splicing activity, especially for A3 events (Fig. 4c,d).** We also observed that DNA methylation in exons and introns of large genes was distinguishable (Fig. 3e and Extended Data Fig. 4e), and genes with low splicing activity were often associated with higher DNA methylation at splice sites. Previous studies showed that TE insertions can alter splicing patterns and even turn off gene expression⁶³, for example through CHG methylation at splice sites in maize⁶⁴. We similarly observed that strong CHG methylation in non-TE regions of introns was associated with low splicing activity in *L. regale* (Extended Data Fig. 5g). This suggested that the mechanisms by which splicing was affected by DNA methylation and TE insertions were **probably different**. Notably, genes with low alternative splicing, are primarily involved in RNA formation and splicing (Supplementary Fig. 8), and contained fewer TE insertions thus relatively stable, likely benefiting from rapid TE clearance. Combination with high CHG methylation at non-TE sites also might reduce splicing variants to stabilize gene transcription. Furthermore, lily-specific gene family analysis indicated that lilies had evolved numerous genes related to RNA splicing, such as poly(A)-specific ribonuclease activity. In conclusion, the extensive TE insertions had contributed to regulating the alternative splicing of genes in *L. regale* who likely evolved specific RNA splicing-related genes (**Supplementary Fig. 8**) to facilitate the proper splicing of the prevalent long genes as a mechanism for environment adaptation."*

Since *L. sargentiae* is not a chromosome-level assembly, we primarily conducted a comparative analysis of *L. regale* with *L. davidii*. The comparative analysis process was as follows: We used OrthoFinder to identify and cluster orthologs across the original 13 genomes and the three newly incorporated genomes from the Liliales order. A phylogenetic tree was then constructed based on 370 low-copy orthogroups to determine species relationships. Next, we analyzed lineage-specific genes in the *Lilium* genus and *L. regale* by examining their clustered orthogroups. We have incorporated the new results into the Results section **Lines 111-125**:

"To understand the evolutionary history of royal lily, we conducted a phylogenomic analysis of 16 species including our L. regale genome and 15 other published genomes including L. davidii²³, L. sargentiae and Gloriosa superba²⁸. Ortholog identification and clustering using Orthofinder³⁵ classified 698,581 genes into 62,451 orthogroups with four categories: multi-copy, single-copy, unique-copy and others (Fig. 1d), 370 of which were composed of low-copy orthogroups (1 or 2 genes per orthogroup). In Lilium, 30,012 unique genes were identified, primarily associated with transcription regulation and post-transcriptional gene silencing (Supplementary Fig. 3). In contrast, L. regale had 15,691 unique genes, which were enriched in nucleotide excision repair and mRNA 3'-UTR binding. These genes may have evolved as adaptations to the exceptionally large genomes of Lilium (Supplementary Fig. 4). A phylogenetic tree constructed using MCMCTREE³⁶ with the 370 low-copy orthogroups, with divergence times estimated and calibrated using fossil records from TimeTrees³⁷ (Extended Data Fig. 2a), showing that G. superba and (L. davidii, (L. regale, L. sargentiae)) diverged at ~95.6 million years ago (Mya). L. regale and L. sargentiae are more closely related than each with L. davidii, consistent with previous transcriptome-based analysis²⁸."

Furthermore, we used CAFE5 to identify significantly expanded orthogroups in the *Lilium* and *L. regale*. Notably, the expanded gene families in *L. regale* were primarily associated with defense pathways. We have incorporated the results into the Results section **Lines 126-130**:

"Furthermore, CAFE³⁸ analysis revealed significantly expanded orthogroups in Lilium primarily involved in terpenoid biosynthesis (Supplementary Fig. 4a). L. regale expanded orthogroups were also enriched with MAPK signaling pathway and defense responses mediated by cell wall thickening (Supplementary Fig. 4b). This reflected the genome evolution in L. regale to cope with environmental stress during its independent evolution^{39,40}."

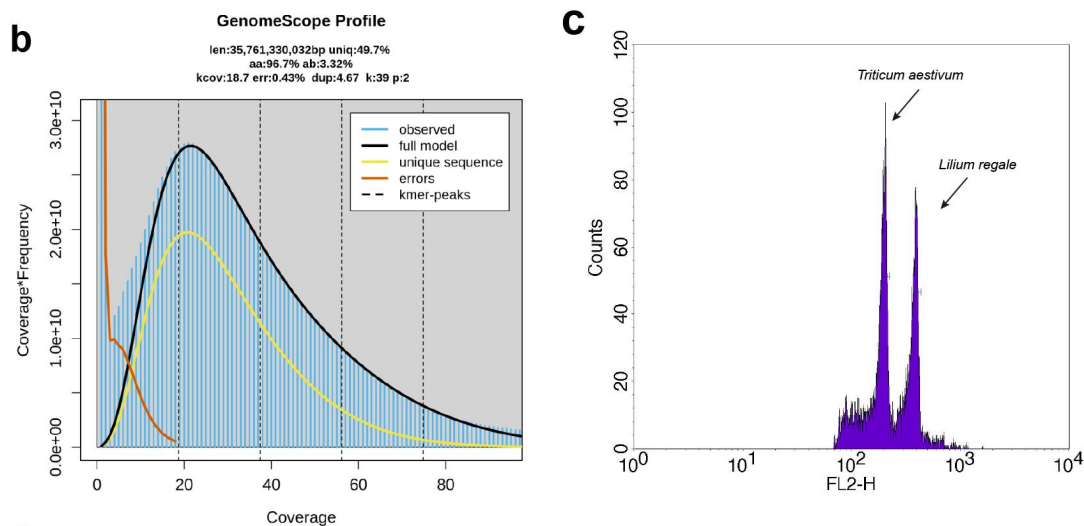
Additionally, based on the analysis of WGD events in *L. davidii*, *L. sargentiae*, and *G. superba*, our findings suggest that the *Lilium* genus underwent a shared whole-genome duplication (WGD) event distinct from that of the Colchicaceae family. We have incorporated this into the Results section **Lines 151-156**:

"Furthermore, synteny analysis between the two lilies L. regale and L. davidii revealed strong gene-level collinearity (Fig. 1g and Extended Data Fig. 2d). Previous studies^{23,28} have shown that L. davidii and L. sargentiae underwent independent WGD events, suggesting that the three Lilium species share a similar WGD history. Together,

these analyses provide robust evidence that *Lilium* species has undergone an independent WGD."

Comment 5. L89: "estimated genome size of 35.8 Gb..." Different values of genome size are given throughout the manuscript - in Supplementary Table 2 this value is 36.12 Gb, in L102 35.6 Gb. I wonder which value is correct? And was the genome size estimated by an independent method, ideally flow cytometry?

Response: Thank you for your comment. We made a mistake about the value of 36.12 Gb in **Supplementary Table 2** and it should be 35.6 Gb. Additionally, we conducted flow cytometry analysis using the bread wheat genome as a reference, which estimated the genome size of *Lilium regale* to be 32.6 Gb (**Supplementary Fig. 1b**). This value is smaller than our draft genome assembly size and *K*-mer-based result, possibly due to the high heterozygosity (3.32%) of the royal lily genome.



Extended Data Figure 1b-c. Estimation of *L. regale* genome size by K-mer (**b**) and flow cytometry analysis (**c**).

Comment 6. L107: This section of the manuscript is about assembly evaluation and refers to Fig. 1c and Extended data Fig. 1c as proof that there are no misassembled blocks or misplaced contigs. I'm not sure how the referenced circos plots prove this.

Response: Apology. There was a figure citation error in this section, and has been corrected in the main text:

Lines 98-100: "We extensively evaluated genome assembly using multiple methods. Hi-C interaction maps showed no apparent large-scale mis-assembly or any misplacement of contigs (**Supplementary Fig. 1,2**)."

Comment 7. L117: "...using RNA-seq.." However, ISOseq data was also used.

Response: Revised as follow:

Lines 104-109: “*Genome annotation by integrating ab initio prediction, homologous proteins, and transcriptome-based evidence from different tissues using ISO-seq and RNA-seq (Supplementary Table 6) predicted 67,862 high-confidence gene models, of which 77.64% were annotated by eggNOG-mapper^{31,32}, Gene Ontology³³ (GO) and Kyoto Encyclopedia of Genes and Genome³⁴ (KEGG), and 52.27% were expressed in at least one tissue (TPM >0).*”

Comment 8. L125: References to the *Allium* genomes included in the comparison are missing,⁰

Response: References to the *Allium* genomes are added. To make the results section more concise, we have moved the references to non-*Lilium* species used for constructing the phylogenetic tree to the methods section.

Line 569-571: “*Orthologues and orthogroups in the 15 monocots^{23,28,44,89,90,97-104} and one dicotyledon¹⁰⁵ as an outgroup (Supplementary Table 16), were inferred using OrthoFinder v2.5.4³⁵ with default values setting and ‘-M msa’ activated.*”

Comment 9. L127-128: The number of orthogroups composed of single-copy orthologs seems rather low to me. Is there an explanation for this?

Response: Thank you for this comment. The low number of single-copy orthologs is mainly because most of the species used in the tree have undergone one or more rounds of WGD events, resulting in most genes having multiple copies. After WGD, gene duplications, typical in plant genomes, may also play a role in this result. After adding the published lily genomes in our analysis, this number gets smaller. Therefore, we revised our analysis by using 370 low-copy orthologs to improve the reliability of the phylogeny (Line 122).

Comment 10. L129: Was the phylogenetic tree based on 24 loci (single-copy genes)?

Response: Yes, the original phylogenetic tree was based on 24 single-copy orthogroups. As explained above, we have revised our analysis by using 370 low-copy orthologs after adding more species in the orthogroup identification including *L. davidii* (Xu, et al., 2024) and other two liliales species (*L. sargentiae* and *Gloriosa superba*).

Comment 11. L131-135: Would the number of unique genes be as high if *L. davidii*

were included in the analysis?

Response: We added the *L. davidii* (Xu, et al., 2024) and other two liliales species (*L. sargentiae* and *Gloriosa superba*) (Liang, et al., 2025). The number of unique genes in *L. regale* decreased from 24,457 to 15,691(**Fig. 1d**).

Lines 118:119 "*In contrast, L. regale had 15,691 unique genes, which were enriched in nucleotide excision repair and mRNA 3'-UTR binding.*"

Comment 12. L172-173: refers to Supplementary Table 2, but the information in the text and that in the table are different; while the text indicates the proportions of LTR retrotransposons of the total TEs, the table indicates the proportions of LTR retrotransposons in the genome. This should be consistent.

Response: Thank you for your comment. We think that you may be referring to **Supplementary Table 7**, which indicates the proportion of LTRs within the entire genome sequence. While in **Lines 170–172**: “*Most of the repetitive were TEs (27.5 Gb) especially 19.8 Gb long terminal repeat retrotransposons (LTR-RTs) taking 72.3% of total TEs (Extended Data Fig. 3a and Supplementary Table 7)*”, it represents the proportion of LTRs within all transposable elements (TEs).

Comment 13. L174-176: This sentence is redundant, gives the same information as the above statement.

Response: Thank you for this comment. We have reformulated the sentences.

Lines 173-178: "*Distinct from humans whose genome contained a higher proportion of LINEs (Long Interspersed Nuclear Elements) and SINEs (Short Interspersed Nuclear Elements), the proportion of Ty3/Gypsy elements increased with genome size in plants (Fig. 2a). However, the fraction of different TE types in L. regale was comparable to smaller genomes such as maize and onion (Supplementary Table 8), suggesting that L. regale genome expansion was caused by amplification of all TE types.*"

Comment 14. L177-180: This sentence is unclear, should be reformulated.

Response: Thank you for this comment. We have reformulated the sentences.

Lines 175-178: “*However, the fraction of different TE types in L. regale was comparable to smaller genomes such as maize and onion (Supplementary Table 8), suggesting that L. regale genome expansion was caused by amplification of all TE*

types."

Comment 15. L182-183: The text refers to Figure 2b and Extended data Figure 3b when mentioning the number of intact Gypsy and Copia elements, but these figures do not show these numbers. In fact, they show the phylogenetic relationships within the two LTR superfamilies.

Response: Thank you for this comment. The redundant citations have been removed.

Lines 178-181: *"Despite a Gypsy-rich genome, the numbers of intact Gypsy (34,472) and Copia (35,771) elements were comparable (Fig. 2b and Extended Data Fig. 3b), with Tekay and Athila as the most abundant LTR/Gypsy subfamilies (Fig. 2b), which was consistent with the research in L. davidii and L. sargentiae^{23,28}."*

Comment 16. L188-189: "...comprising 3.28%, 19.63%, and 49.20% of the total LTR content... (Supplementary Table 9)" Again, the table has different numbers. This should be consistent.

Response: Thank you a lot for this comment. We have corrected the number to make it consistent.

Lines 183-186: *"A total of 79,821 intact LTRs (650.88 Mb), 4,249,484 truncated LTRs (7.13 Gb), and 19,230,536 solo LTRs (9.11 Gb) were identified in L. regale, comprising 3.28%, 35.90%, and 45.88% of the total LTR content (19,858.9 Mb), respectively (Supplementary Table 9), suggesting the rapid recombination of LTRs. "*

Comment 17. L190-192: The oldest LTR insertion is dated to 6.39 Mya, but this value is not visible as a distinct peak in Fig. 2c, so it has no relevance for genome size expansions due to increased LTR activity.

Response: Thank you for this comment. We have removed the redundant description and revised the text as below to make it clear.

Lines 186-189: *"All chromosomes exhibited high density peaks especially towards both ends indicating LTR-RT hotspots (Fig. 2c), mainly composed of dense insertions from the same LTR lineage within a short time frame."*

Comment 18. L210-211: The reported stronger H3K9Me2 enrichment at the termini of the intact LTRs is not obvious from Fig. 2e, which is referenced to demonstrate it.

Response: Thank you for this comment. Sorry for the confusion. We intended to convey that the intact LTR regions exhibit more pronounced H3K9Me2 enrichment near both

ends compared to the protein regions in the middle, as shown by the peaks circled with red dashed lines in **Fig. 2e**. We also have revised the description in the result section.

Lines 198-201: *"We examined the H3K9Me2 marks across different types of LTR regions and their flanking sequences, and observed a significant higher enrichment of H3K9Me2 in solo-LTRs than truncated LTRs while intact LTRs showed higher enrichment near terminals than middle (Fig. 2e),"*

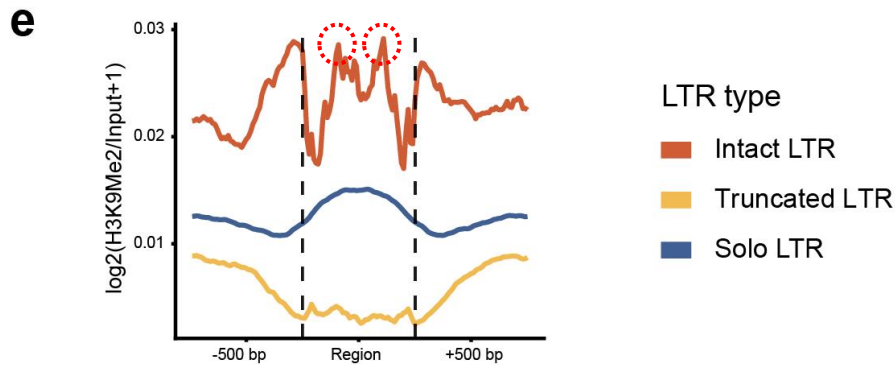


Figure 2e. Enrichment of H3K9Me2 on different types of LTRs.

Comment 19. L214-215: "However, the truncated LTR resulting from this insertion..." What information could be used to infer the insertion time of a truncated LTR retrotransposon (i.e., an LTR retrotransposon that does not contain terminal repeats)?

Response: Thank you for your comment. We did not identify the exact insertion time of truncated and solo LTRs here, only by finding their homolog intact LTR as queries in sequence alignment (BLAST), applying the criteria of identity >80% and e-value <1e-4. And it also has been implied in the methods section.

Lines 544-545: *"The homolog truncated and solo LTRs by using intact LTRs as queries in sequence alignment (BLAST), applying the criteria of identity >80% and e-value <1e-4."*

Comment 20. L262-263: The reported differences in TE insertions and DNA methylation between introns and exons are not obvious from Figure 3f. The exons and introns should be clearly labeled in the figure.

Response: Thank you for this comment. We have marked the exon regions by grey shades to make it clearer.

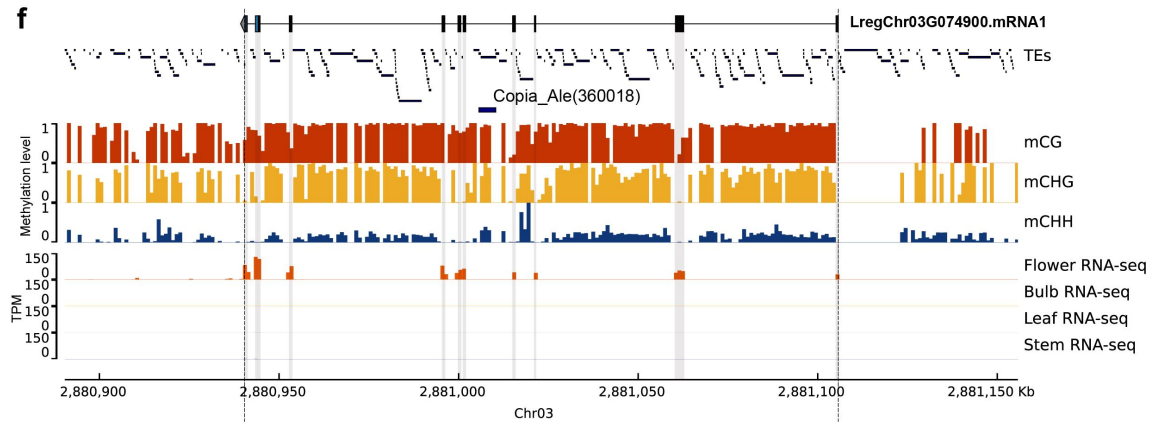


Figure 3f. with many TE insertions, *LregChr03G072890* expression under the control of methylation. The gene structure, TE insertions, intact LTR with its insertion time (year), methylation levels (CG, red; CHG, yellow; and CHH, blue; methylation level represent the proportion of 5mC at the given site), gene expression in four tissues (flower, orange; bulb, canary; leaf, cyan; and stem, purple) were plotted from the top to bottom successively. The shaded gray areas represent exon regions.

Comment 21. L296-297: I don't see the described patterns in the referenced figures.

Response: Thank you for your comment. As shown in **Fig. 4e**, DEPGs (difference psi genes) exhibit minimal differences in DNA methylation levels within exons. However, there are substantial differences in CHG site methylation levels within introns. Specifically, LPGs display higher CHG methylation levels (indicated by yellow dashed lines) compared to the average levels (gray solid lines), and this trend is consistent across different tissues (**Extended Data Fig. 6e**).

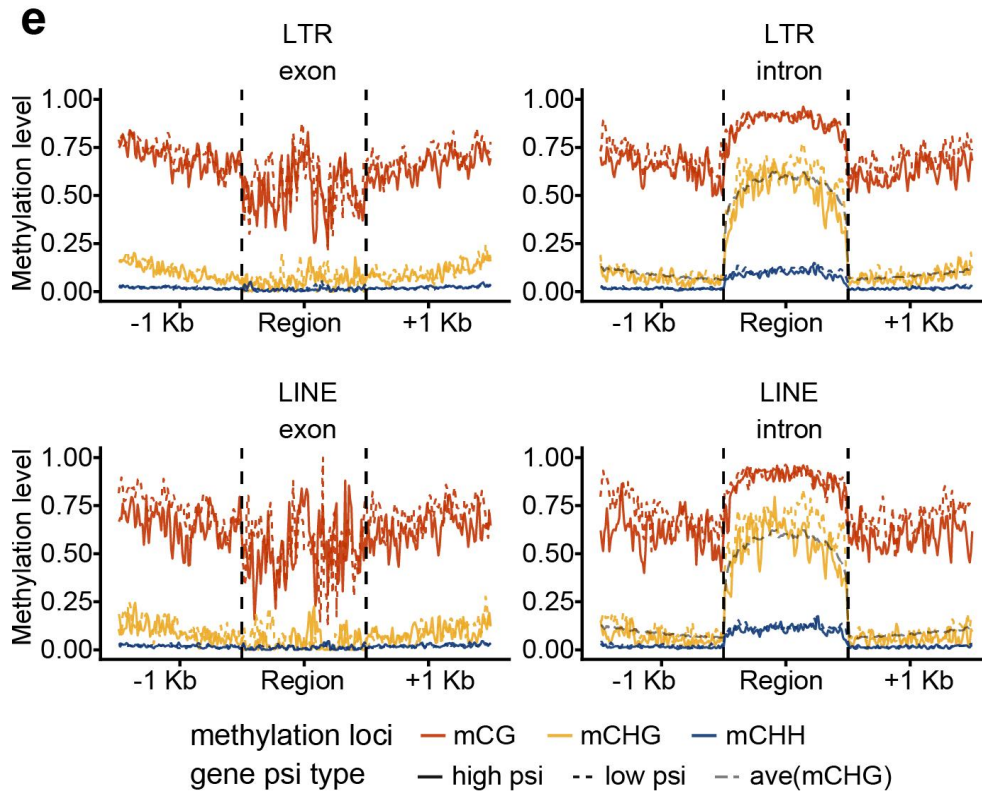
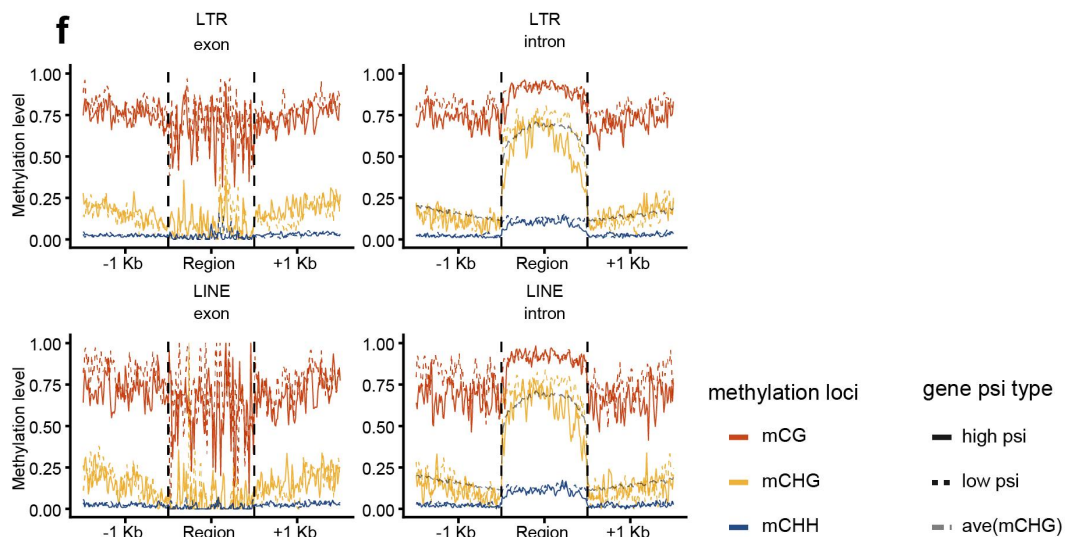


Figure 4e. The DNA methylation level (CG, red; CHG, yellow; CHH, blue; methylation level represents the proportion of 5mC at the given site) of the gene bodies, upstream and downstream of genes with LTR/LINE insertions (colored line, HPGs; colored dash, LPGs; grey dash, average of all genes at mCHG).



Extended Data Figure 6f. The methylation levels of exons and introns for different PSI genes with LTR and LINE insertions in leaves.

Comment 22. L367-368: "...most of the TPR genes involved in RNA splicing were

unique to *L. regale*, suggesting the presence of a unique splicing mechanism in *L. regale*..." I wonder if this mechanism is really unique to the royal lily or is it more genus or family specific? Again, this could be answered by including the *L. davidii* genome in the analysis.

Response: Thank you for your insightful comment. Thank you for your insightful comment. We have also conducted functional enrichment of unique genes from each of the three additional genomes within the Liliales order: *Lilium davidii*, *Lilium sargentiae*, *Lilium regale*, and *Gloriosa superba*. Interestingly, only in *L. regale* and *G. superba* did we find significant enrichment of the unique genes in splicing-related mechanism, namely the *CAF1* gene family (CCR4 associated factor 1, has known function of mRNA Poly(A) tail shortening regulation, a key process in A3 type alternative splicing events) (Liang et al., 2009, and Fang et al., 2021). However, due to lack of TE annotation for the other three genomes (not provided in their publications), we are unable to determine whether this TE-density regulatory mechanism of splicing is unique to *L. regale* or prevalent across the Liliales family. Considering this limitation, we cannot draw a solid conclusion about whether this is unique mechanism for *L. regale* without further experimentation. Therefore, we have removed the previous statement about the "unique mechanism".

References:

Liang W, Li C, Liu F, Jiang H, Li S, Sun J, Wu X, Li C. The Arabidopsis homologs of CCR4-associated factor 1 show mRNA deadenylation activity and play a role in plant defence responses. *Cell Res.* 2009 Mar;19(3):307-16.

Fang JC, Tsai YC, Chou WL, Liu HY, Chang CC, Wu SJ, Lu CA. A CCR4-associated factor 1, *OsCAF1B*, confers tolerance of low-temperature stress to rice seedlings. *Plant Mol Biol.* 2021 Jan;105(1-2):177-192.

Comment 23. L604: What script was used to identify the solo LTRs?

Response: Thank you for this comment. solo LTRs were identified by the script which is provided by the author of LTR_retriever, and you can find it at:

https://github.com/oushujun/LTR_retriever/blob/master/bin/solo_finder.pl.

The information has also been added to methods. **Lines 542-544:** "*Solo-LTRs were identified using the script supplied by LTR Retriever; then the potential solo-LTRs were classified by TESorter with truncated-LTRs and other TEs removed using bedtools*

v2.30.0⁸⁶."

Comment 24. L612-614: Why were the protein data from *L. davidii* not used for homology-based gene identification along with the others mentioned?

Response: Thank you for your comment. At the time of completing our work, *Lilium davidii* genome had not yet been published, so we did not use their homologous proteins for annotation. However, based on the BUSCO results for protein annotation, the quality of our gene annotations (88.20%) is better than *L. davidii* (79.50%). Additionally, we performed orthologous annotation of proteins from three lilies using Orthovenn3 (Sun et al., 2023), resulting in 13,270 shared clusters. This indicates that a substantial number of homologous proteins were accurately annotated (Fig. R1).

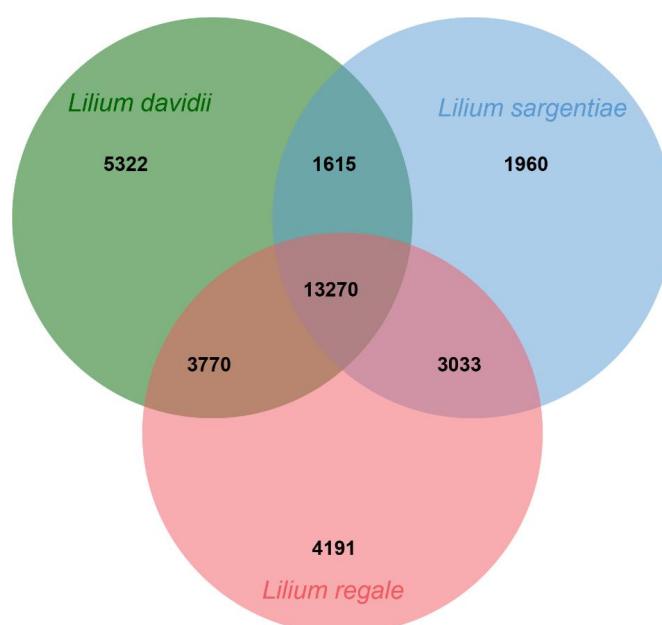


Fig. R1. The specific or shared orthologous between three lily genomes

Reference:

Sun J, Lu F, Luo Y, et al. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. Nucleic Acids Res. 2023 Jul 5;51(W1):W397-W403.

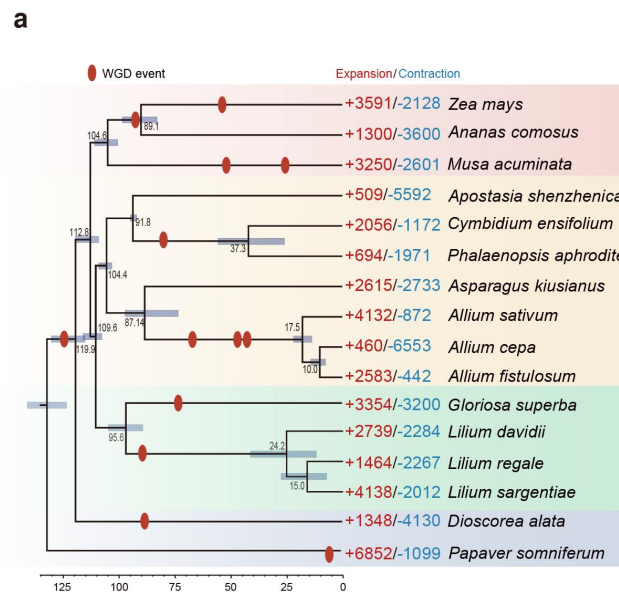
Comment 25. L637-641: divergence time dating was performed, but the results of this analysis are not included in the manuscript.

Response: Thank you for this comment. We sincerely apologize for the mistake. We have added the description of divergence time dating in the results section and incorporated the divergence time into the phylogenetic tree (**Extended Data Fig. 2a**).

Lines 121-125: "A phylogenetic tree constructed using MCMCTREE³⁶ with the 370 low-copy orthogroups, with divergence times estimated and calibrated using fossil records from TimeTrees³⁷ (Extended Data Fig. 2a), showing that *G. superba* and (*L. davidii*, (*L. regale*, *L. sargentiae*)) diverged at ~95.6 million years ago (Mya). *L. regale* and *L. sargentiae* are more closely related than each with *L. davidii*, consistent with previous transcriptome-based analysis²⁸. "

References:

Liang Y, Gao Q, Li F, et al. The giant genome of lily provides insights into the hybridization of cultivated lilies. *Nat Commun.* 2025 Jan 2;16(1):45.



Extended Data Fig. 2a. A maximum likelihood phylogenetic tree of the genomes of 16 species. The numbers near the internal nodes represent the estimated divergence time, while the blue bars indicate the 95% confidence interval for the estimated divergence time. The number of gene-family expansion and contraction events ($P \leq 0.01$) is depicted in red and blue, respectively.

Response to Reviewer comments

Reviewer #1 (Remarks to the Author):

My concerns have been properly addressed and this manuscript was comprehensively improved. The Royal Lily genome could be a valuable resource for plant community.

RESPONSE: Thank you very much for your positive comments!

Reviewer #2 (Remarks to the Author):

I have reviewed the revised version of the manuscript. The authors have addressed all the comments I raised in my previous review. They shortened the results section, improved the clarity and informational value of the figure captions, and provided more detailed explanations of some of the results. Moreover, they included additional *Lilium* genomes and reworked the comparative genomic analyses, enabling the distinction of patterns specific to *L. regale* from those shared with other *Lilium* species. I have no further comments on the manuscript and recommend it for publication.

RESPONSE: Thank you very much for your positive comments!